# Methods

The Coalition for Improving Maternity Services:

Henci Goer, BA

**ABSTRACT**

In this article, the details of the methods used to determine the evidence basis of the *Ten Steps of Mother-Friendly Care* are presented and discussed.

*Journal of Perinatal Education*, *16*(1–Supplement), 5S–9S, doi: 10.1624/105812407X173128
*Keywords:* systematic review, Ten Steps of Mother-Friendly Care, Agency for Healthcare Research and Quality standards

The systematic review is the optimal vehicle for establishing a detailed evidence basis for the *Ten Steps of Mother-Friendly Care,* developed by the Coalition for Improving Maternity Services (CIMS). Ebell et al. (2004) define a systematic review as "a critical assessment of existing evidence that addresses a focused clinical question, includes a comprehensive literature search, appraises the quality of studies, and reports results in a systematic manner" (p. 549). This process gives systematic reviews important advantages over the more conventional, narrative review, as described (Goer, in press):

- Systematic reviews cast a wide net. With narrative reviews, no attempt is made to retrieve all relevant research; instead, reviewers choose what suits their thesis.
- Systematic reviews describe their methodology. Narrative reviews make explicit neither how reviewers went about selecting studies nor the basis on which studies were included or excluded.
- Systematic reviews apply uniform criteria. Narrative reviewers may include or reject a study simply because they like or do not like its conclusions.
- Systematic reviews evaluate quality. Narrative reviews behave as if all studies are alike, whereas systematic reviews include only higher quality studies. This means that, unlike narrative reviews, systematic reviews draw conclusions from the best evidence available. Systematic reviews also clarify where there is insufficient evidence to reach a conclusion.
- Systematic reviews report results in a structured way. Narrative reviews tend to cite specific results from a few studies in support of a theory.

It would seem at first glance that a valid systematic review would not be possible given that the *Ten Steps of Mother-Friendly Care*, the conclusions of the proposed review, were already fixed. However, when the *Ten Steps* were developed, only those steps for which research had established consensus or which were intuitively obvious as "best practice" were included. The task for this project, therefore, was refined to evaluate and present the quality of evidence supporting specific rationales for each of the *Ten Steps*.

The review expanded on conventional systematic reviews in that it addressed a broad range of

*For more information on the Coalition for Improving Maternity Services (CIMS) and copies of the* Mother-Friendly Childbirth Initiative *and accompanying* Ten Steps of Mother-Friendly Care, *log on to the organization's Web site (www.motherfriendly.org) or call CIMS toll-free at 888-282-2467.*

outcomes of interest. The content of conventional reviews are generally confined to the presence or absence of short-term, adverse, physical outcomes. They also typically evaluate the use of specific interventions in isolation rather than considering the effects of a "high-intervention" system of care versus one that is not.

Members of the CIMS Expert Work Group (EWG) conducting this review, like the developers of the *Mother-Friendly Childbirth Initiative* itself, recognized that the absence of disease does not equal health. They also recognized that the excessive use of intervention is, in itself, harmful because it imposes risks with no evidence of benefit. Accordingly, the EWG examined long-term outcomes, psychosocial outcomes, quality of life concerns, the impact of birth practices on breastfeeding, increased need for further medical intervention, and short-term morbidity.


*Members of the CIMS Expert Work Group were:*

- *Henci Goer, BA, Project Director*
- *Mayri Sagady Leslie, MSN, CNM*
- *Judith Lothian, PhD, RN, LCCE, FACCE*
- *Amy Romano, MSN, CNM*
- *Karen Salt, CCE, MA*
- *Katherine Shealy, MPH, IBCLC, RLC*
- *Sharon Storton, MA, CCHT, LMFT*
- *Deborah Woolley, PhD, CNM, LCCE*


## PROJECT DESIGN

The EWG consisted of eight people. Members came from varied professional backgrounds, were committed to mother-friendly care, and were knowledgeable about either maternity care research in general or the research in their specific field. EWG members had expertise in the various aspects of mother-friendly care covering all elements of the *Ten Steps of Mother-Friendly Care*.

The *Ten Steps* were parceled out among six members of the EWG for research and review (HG, MSL, KS, KS, SS, DW). In accordance with the requirements of systematic reviews, EWG members determined whether to include or exclude studies based on specific criteria (see later discussion). They extracted data from each included study into a data summary sheet and listed a reason for each study they excluded. The EWG developed the data summary template based on guidelines published by the Agency for Healthcare Research and Quality (AHRQ) and an article recommending strategies for conducting valid systematic reviews with limited resources (Griffiths, 2002; West et al., 2002).

To provide intra- and interobserver reliability, one member of the EWG who did not participate in the primary review process served as a "second reader" (AR). The second reader and project director (HG) determined which topics would require a second reading. The topics chosen represented the steps (or components thereof) that were considered most controversial in the literature and/or in practice and included the following: home birth, freestanding birth centers, routine intravenous lines, withholding food and drink in labor, routine early amniotomy, routine electronic fetal monitoring (cardiotocography), induction rate, cesarean-section rate, vaginal birth after cesarean rate, hydrotherapy, epidurals, circumcision, and adoption of baby-friendly status. The second reader was then responsible for reading and independently evaluating the quality of the studies that were reviewed for the preselected topics and reviewing all data summary sheets to ensure they were correct and complete. Finally, with no knowledge of the rating assigned by the primary reviewers, the reader assigned ratings of the strength of the aggregate evidence supporting each rationale for the three domains (see later discussion). Any discrepancies between the ratings assigned by the primary reviewer and the second reader were resolved by consensus. Another EWG member (JL) assumed the role of project director during the final stages of the process and was involved with writing, editing, and preparing the document for publication.

## DATA SOURCES

EWG members conducted searches in the following seven databases: CINAHL, the Cochrane Library, DARE, MEDLINE, OMNI, PsychINFO, and Scirus. In addition, EWG members obtained studies from their own files and the reference lists of other studies and reviews (both narrative and systematic). EWG members included studies published between January 1, 1990, and June 1, 2006.

## EXCLUSION CRITERIA

Study exclusions came in two categories: absolute and relative. Absolute exclusions were the following:

- Studies published in languages other than English. Fortunately, many studies carried out in countries where English is not the native language are published in English-language journals.
- Studies available only as an abstract.
- Narrative reviews, commentaries, or practice guidelines. Narrative reviews and commentaries are opinion pieces. Opinion pieces are the weakest form of evidence and were disqualified on that basis. Practice guidelines are generally opinion pieces as well, but even where they are not, unlike systematic reviews, they do not provide the information necessary to evaluate the quality of the literature review on which they are based.
- Studies with surrogate outcomes, with two exceptions (see later discussion).

Grimes and Schulz (2005) define a surrogate outcome, also called "surrogate marker" or "intermediate measure," as "an outcome measure that substitutes for a clinical event of true importance" (p. 1114). Surrogate outcomes are usually laboratory or imaging studies "thought to be in the causal pathway to a clinical event of interest" (p. 1114). For example, a measurement of pelvic-floor muscle strength or an ultrasound scan showing a defect in the anal sphincter muscle would be surrogate outcomes as opposed to outcomes measuring urinary or bowel incontinence. Surrogate outcomes often correlate poorly with clinical outcomes, as is the case with the examples cited here. Nonetheless, although surrogate outcomes cannot rule in adverse clinical outcomes, they can sometimes be useful in ruling them out. Using the current example, the fact that the pelvic-floor musculature is stronger in women who have spontaneous tears defeats the argument that episiotomy prevents urinary incontinence.

The second situation in which surrogate outcomes can be useful is in cases whereby the endpoint is so rare that it is not feasible to conduct a study large enough—that is, with sufficient statistical power—to have a reasonable chance to detect differences between groups. Neonatal death in full-term pregnancies with no medical complications serves as one example. In such cases, studies relying on surrogate outcomes were acceptable, provided the outcome was closely linked to the actual event of interest and could be measured objectively, as when newborns required prolonged respiratory assistance as opposed to low 5-minute APGAR scores.

Relative exclusions were left to the individual judgment of the EWG member and depended on the specific topic being researched. Relative exclusions were the following:

- On rare occasions and for reasons listed with the reference, studies published earlier than 1990.
- Studies in countries lacking medical resources.
- Weaker studies (see later discussion for grading scheme).
- Studies included systematic reviews.
- Multiple reports on the same study or dataset.

Studies published more than 15 years prior to conducting this review or published in countries lacking medical resources were excluded to ensure that results could be generalized to modern medical care. Nonetheless, outcomes of interest might not depend on these factors, and what constitutes a weaker study varies from rationale to rationale, depending on what evidence is available.

Individual studies analyzed in systematic reviews were excluded to avoid duplication. Exceptions were made for the rare case in which the systematic review did not report an outcome of interest, but individual studies included in the review did. It should be noted, however, that systematic reviews often overlapped in the studies they included. As for multiple reports on the same study or dataset, only those reports containing unique data pertinent to the rationales for each of the *Ten Steps of Mother-Friendly Care* are cited.

Finally, the EWG took into account the degree to which protocol was violated in randomized controlled trials. Randomized controlled trials are analyzed according to "intent to treat," not actual treatment, because to do otherwise defeats the purpose of random assignment. If a few participants receive the treatment of another group, this is not a problem; but in obstetric trials, it is not uncommon for sizeable percentages to be given the treatment of another group. This crossover decreases the power of the trial to detect differences between groups. For example, investigators conducting a randomized controlled trial of epidural analgesia versus nonepidural pain relief calculated that 263 women per group would be needed to have an 80% probability of detecting a doubling of the cesarean rate from 7% to 15%, assuming that the noncompliance rates were 25% to 30% in the nonepidural group (Dickinson, Paech, McDonald, & Evans, 2002). The actual noncompliance rate was 60%, which would require 12,000 participants to detect the same difference. In some cases, trials and reviews were excluded on this basis; but in others, it was not feasible to do so. Explanatory notes alert the reader to this caveat, where relevant.

## GRADING SCHEME

Individual studies were given a quality rating using guidelines published by the AHRQ (West et al., 2002). The following selected elements recommended by AHRQ were considered when evaluating individual studies and, on this basis, each included study was graded good, fair, or weak:

### *Systematic Reviews*

- Study question: "question clearly specified and appropriate."
- Search strategy: "sufficiently comprehensive and rigorous."

- Inclusion and exclusion criteria: ''selection methods specified and appropriate.''
- Data extraction: ''rigor and consistency of process,'' ''measure of agreement or reproducibility'' [*Note:* This only applies to reviews that include meta-analyses], ''extraction of clearly defined interventions/exposures and outcomes for all relevant subjects and subgroups.''
- Study quality and validity: ''assessment method specified and appropriate.''
- Data synthesis and analysis: ''appropriate use of qualitative and/or quantitative synthesis, with consideration of the robustness of results and heterogeneity issues.''
- Funding or sponsorship: ''type and sources of support for study.''

***Randomized Controlled Trials***

- Study question: ''clearly focused and appropriate question.''
- Study population: ''specific inclusion and exclusion criteria.''
- Randomization: ''adequate concealment method used.''
- Blinding: ''double-blinding (e.g., of investigators, caregivers, subjects, assessors, and other key study personnel as appropriate) to treatment allocation.''
- Interventions: ''intervention(s) clearly detailed for all study groups.''
- Outcomes: ''primary and secondary outcome measures specified;'' ''assessment method standard, valid, and reliable.''
- Statistical analysis: ''appropriate analytic techniques that address study withdrawals, loss to follow-up, missing data, and intention to treat;'' ''power calculation;'' ''assessment of confounding [factors].''
- Results: ''measure of effect for outcomes and appropriate measure of precision.''
- Funding or sponsorship: ''type and sources of support for study.''

***Observational Studies***

- Study question: ''clearly focused and appropriate question.''
- Study population: ''description of study populations.''
- Comparability of participants: ''specific inclusion/exclusion criteria for all groups,'' ''criteria applied equally to all groups,'' ''comparability of groups at baseline,'' ''comparability of follow-up among groups at each assessment,'' ''explicit case definition [case–control studies],'' ''controls similar to cases except without condition of interest and with equal opportunity for exposure [case–control studies].''
- Exposure or intervention: ''clear definition of exposure;'' ''measurement method standard, valid and reliable;'' ''exposure measured equally in all study groups.''
- Outcome measurement: ''primary/secondary outcomes clearly defined;'' ''outcomes assessed blind to exposure or intervention status;'' ''method of outcome assessment standard, valid and reliable;'' ''length of follow-up adequate for question.''
- Statistical analysis: ''power calculation provided,'' ''assessment of confounding [factors].''
- Results: ''measure of effect for outcomes and appropriate measure of precision.''
- Funding or sponsorship: ''types and sources of support for study.''

Also using AHRQ's precepts, the strength of the aggregate evidence supporting each rationale was graded A, B, or C in three domains (West et al., 2002):

- Quality: ''the aggregate of quality ratings for individual studies.''
- Quantity: ''magnitude of effect, numbers of studies, and sample size or power.''
- Consistency: ''the extent to which similar findings are reported using similar and different study designs.''

Because these domains function independently of each other, they provide a more nuanced evaluation than the usual single-score grading systems. This system also corrects a weakness of systematic reviews. It makes clear, in contrast to systematic review abstracts, cases where only one study reports on a particular outcome or where the quantity of evidence is small.

EWG members varied somewhat in how they presented their data. As a result, some tables use the term ''may'' versus ''can'' to indicate rationales for which studies disagreed versus those for which studies were consistent.

***Additional Grading Information***

To the AHRQ scheme, the EWG added ''no evidence of benefit'' and ''no evidence of harm.'' The concept of no evidence of benefit was needed for routine interventions (e.g., routine IV drip)

whereby the quality of research could not be ascertained because no research had examined the policy. In these cases, because benefit has not been established but harm has, the policy should be abandoned until such time as research establishes that benefits outweigh the hazards. The concept of no evidence of harm was needed for mother-friendly practices (e.g., freedom of movement during first-stage labor or the companionship of family and friends) for which research has not established benefit other than that women prefer it.

Grading schemes frequently use a hierarchy, placing systematic reviews of randomized trials at the pinnacle followed by individual randomized controlled trials, systematic reviews of observational studies, individual observational studies, and case reports or series. The Oxford Centre for Evidence-Based Medicine is a well-respected example of this approach (Centre for Evidence-Based Medicine, 2001). However, as Glasziou, Vandenbroucke, and Chalmers (2004) point out, different questions require different study types. For example, randomized controlled trials, even in the aggregate, rarely have the power to detect differences in rare, catastrophic outcomes, a category of great importance when exposing healthy women and babies to routine or frequent use of intervention. The EWG, therefore, decided not to give precedence to any study design, with the exception of systematic reviews. Because of their nature, systematic reviews potentially offer the strongest evidence—provided their component studies are sound—because they aggregate evidence from multiple studies. Before including a systematic review, EWG members evaluated the component studies, or at least the larger studies, if the studies were too numerous to make it feasible to evaluate them all. When a systematic review was available on a particular topic, EWG members included it over studies of that same topic published during the time period covered by the review and added qualified studies published subsequent to the review.

## CONCLUSION

Developing a systematic review of the *Ten Steps of Mother-Friendly Care* posed a unique challenge: Medical studies are designed to determine the best ways of predicting, diagnosing, and treating disease. The questions they ask are almost always illness-oriented and take the limited form: ''Which is better: A or B?'' Systematic reviews of medical studies, therefore, have evolved as a means of evaluating bodies of such research.

In contrast, this project evaluated a *system* of care intended to promote health and well-being during a fundamentally normal process. These differences necessarily required adapting the conventional techniques used in systematic reviews while adhering to their basic precepts. In this sense, this review is both an extension and reflection of the *Mother-Friendly Childbirth Initiative*, which itself expanded on conventional strategies for developing practice guidelines. CIMS hopes that the process that resulted in the *Ten Steps of Mother-Friendly Care* and the methodology of this systematic review will serve as models and guidelines for others who wish to base maternity care—indeed, medical care in general—on humanistic, holistic, and egalitarian principles while maintaining scientific rigor.

## REFERENCES


Centre for Evidence-Based Medicine. (2001). *Levels of evidence and grades of recommendation.* Retrieved December 10, 2006, from http://www.cebm.net/levels_of_evidence.asp

Dickinson, J. E., Paech, M. J., McDonald, S. J., & Evans, S. F. (2002). The impact of intrapartum analgesia on labour and delivery outcomes in nulliparous women. *The Australian and New Zealand Journal of Obstetrics and Gynaecology, 42*(1), 59–66.

Ebell, M. H., Siwek, J., Weiss, B. D., Woolf, S. H., Susman, J., Ewigman, B., et al. (2004). Strength of recommendation taxonomy (SORT): A patient-centered approach to grading evidence in the medical literature. *American Family Physician, 69*(3), 548–556.

Glasziou, P., Vandenbroucke, J. P., & Chalmers, I. (2004). Assessing the quality of research. *BMJ, 328*(7430), 39–41.

Goer, H. (in press). *Obstetric myths versus research realities* (2nd ed.). Ann Arbor, MI: University of Michigan Press.

Griffiths, P. (2002). Evidence informing practice: Introducing the mini-review. *British Journal of Community Nursing, 7*(1), 38–39.

Grimes, D. A., & Schulz, K. F. (2005). Surrogate end points in clinical research: Hazardous to your health. *Obstetrics & Gynecology, 105*(5), 1114–1118.

West, S., King, V., Carey, T. S., Lohr, K. N., McKoy, N., Sutton, S. F., et al. (2002). *Systems to rate the strength of scientific evidence* (No. AHRQ Publication 02-E016). Rockville, MD: Agency for Healthcare Research and Quality.


---


HENCI GOER *is an award-winning medical writer, author of* The Thinking Woman's Guide to a Better Birth *and* Obstetric Myths and Research Realities*, and an internationally known speaker. An independent scholar, Goer is an acknowledged expert on evidence-based maternity care. She is currently a resident expert for the Lamaze Institute for Normal Birth and moderates the online* Normal Birth Forum *(www.normalbirth.lamaze.org).*